DEVELOPMENT AND USE OF A GEOGRAPHIC INFORMATION SYSTEM (GIS) FOR RESOURCE APPRAISAL

By W. R. Beeman

A geographic information system (GIS) is a computer program that allows one to access or display data spatially. A computer program that can locate points, lines or areas geographically, associate data with them, and allow retrievals and calculations to be made based upon geographic locations is a GIS. GIS programs can vary in complexity from simple, in-house programs that are used to perform basic calculations, to large commercial packages that may perform a myriad of functions. Most of the GIS work for this assessment was done using a commercial package known as ARC/INFO, which is a product of Environmental Systems Research Institute (ESRI). The basic unit in a GIS is called a coverage and is the computer equivalent of a paper map. Coverages can be displayed as backgrounds on computer screens, used as a means for retrieving or assigning values to data points, or be plotted as hard-copy paper maps.

The use of GIS for the assessment fell within two broad categories: (1) the display of complex data sets in such a way that they could be comprehended through the use of maps; and (2) the actual manipulation of spatial data in order to perform complex calculations.

A BRIEF HISTORY OF THE DEVELOPMENT OF GIS TECHNIQUES FOR THE ASSESSMENT

The first steps toward utilizing a GIS in assessing petroleum resources were taken during the previous assessment. Richard Mast and David Root designed a FORTRAN program that generalized data from the Well History Control System (WHCS), a commercial database of oil and gas well information compiled by Petroleum Information Corporation. Areas of interest were divided into a grid consisting of square mile cells, and calculations were made to determine whether or not any wells existed in the cells, whether any wells had produced petroleum from a specified stratigraphic interval, the depths of wells in the cells, and the years of completion and initial production. When plans were made to use GIS as a part of the current assessment, it was a logical beginning to use the program developed during the preceding assessment. However, several alterations were made to the program. The size of the cells in the grid were changed to one-quarter mile by one-quarter mile, or 40 acres each. It was also decided to attach field names to any cell where field names existed in the WHCS database. Those cells were then utilized as the primary database for ensuing calculations. A program was developed using Arc Macro Language (AML) and AWK (a non-compiled language available on UNIX-based computer systems) to convert the retrieved data into an ARC/INFO format for display and further calculations.

In order to make the data retrievals it was necessary to define the stratigraphic interval in which petroleum production might occur for each play. Because that is information that could only be determined by careful study of the geology of an area, it was necessary for each individual province geologist to provide the stratigraphic definition of each play. To facilitate the process an inventory was made of producing formations for each province. The province geologist then assigned formations to each play. Each geologist also provided a map that showed the areal extent of each play. The map was then scanned and converted into a coverage in ARC/INFO. Plots were generated showing the play outlines with the calculated cell coverages. Those plots allowed one to see where production occurred within specified stratigraphic intervals, as well as where wells had penetrated the interval but no production was indicated. By using those plots the province geologists were able to reconcile the play outlines with the areas of production and exploration for the defined stratigraphic interval. Final play outlines were then provided by the province geologists, and were scanned and converted into ARC/INFO coverages in the database. Play outlines in a GIS can be used for further data capture and calculations. They were used in conjunction with the cell coverages to determine drilling success ratios within prescribed areas. They were used to set the areal extent of data files as part of a discovery-process model. They were used for illustrative purposes in the final report. Each of these processes had a separate function and used a separate set of programs written in AML.

As the assessment developed, it was determined that it would be helpful to build a program that would utilize the database to automatically generate outlines for oil and gas fields. Although that program was built and did generate the outlines, inconsistencies in naming conventions for established fields, wells that are misidentified or mislocated in the original records, as well as other difficulties, made it necessary for a geologist to review the outlines and make any necessary alterations or otherwise exercise editorial judgment. In order to avoid having that person learn a new programming language, a GUI (graphical user interface) was utilized to allow "point and click" operation of the program. This led to a menu-driven suite of programs that allows one to run any of a number of programs developed for the assessment, or to enter a separate suite of programs to edit the field outlines. A serendipitous result was that the program served as a check on the accuracy of data sets provided by vendors.

A BRIEF TECHNICAL DISCUSSION

When the assessment began it was anticipated that there would be somewhere between one thousand and two thousand GIS coverages that would have to be generated and stored for the final product. There also had to be provisions for works in progress, special coverages to provide backgrounds, geologic or structural coverages for reference, graphics or plotfiles, as well as the many programs that would be developed and run. It was necessary to make the naming conventions and directory structure simple enough that anyone could find a necessary coverage should the need arise. It was also necessary to structure any databases in such a way that categories could be understood and easily accessed, yet storage requirements would be minimized. A directory structure was developed that had four major areas: (1) works in progress; (2) finished products; (3) supporting or background coverages; and, (4) programs. Coverages were assigned a four-digit number, with the first two digits consisting of the province number and the second two digits representing the play number within the province. By assigning a number to each play, and by storing the plays in a directory structure based on those numbers, a program was developed that could automatically find any given play within the directory structure. That program was inserted at the beginning of every program that needed to access play coverages, and allowed individuals with little or no training in programming to perform intricate routines, such as plotting exploration history maps, with a single command.

Four main programming languages were used to perform GIS functions during the assessment: (1) FORTRAN; (2) SQL (Structured Query Language); (3) AML; (4) AWK. The source data sets were stored on a VAX system, using INGRES as a database management system. Data were retrieved from that system using programs written in SQL. FORTRAN was used to calculate grid cells and their attributes and to perform mathematical calculations for the discovery-process model. AML, which is a programming language designed to allow ARC/INFO commands to be bundled into programs, was used to develop plotting routines for the display of coverages, to provide menu-driven "point and click" graphical user interfaces, to perform routines for the assignment of wells and cells to particular plays, and to provide an automated procedure for building field outlines. AWK was used to convert data from various other formats into a form that was compatible with the assessment's database. AWK had the dual advantages of being much more compact than C or FORTRAN and it could be embedded in the AML programs--it was treated by the program as another AML command rather than having to be executed from within the program by a call function.

When making plots of data to display petroleum production there are several issues to be addressed in terms of level of detail. There may be several hundred wells drilled into a one-square-mile area. If all those points are displayed on a single map, there is some scale at which they become an indistinguishable blur. If different symbols are used for production and unsuccessful exploration, there is some scale at which symbols overplot each other or colors appear to run together, rendering the plot less useful. If one tries to overcome the above problems by scaling the symbols for individual wells, there is a point at which the resolution of the coverage is greater than the ability of the plotter to display it. The solution was to generalize the data to a scale that would overcome those difficulties, yet present the data at sufficient resolution to aid in understanding the geology. It was determined early in the assessment that the ideal scale for the presentation of data on paper maps would be 1:500,000. At the same time, the data would eventually be plotted on a national map at 1:3,750,000. A series of plots were generated at various scales using many different symbol sets, and it was determined that the ideal resolution was obtained by generalizing the data to a 40-acre spacing (one-quarter mile by one-quarter mile grid).

Hard copies of the data had to be generated in order for individual province geologists to examine how well the data fit their geographic definition of each play, and hard copies had to be regenerated when necessary to incorporate changing definitions as the assessment progressed and additional input modified each geologist's conception of plays within a province. It was necessary to write programs that minimized the use of personnel to generate the paper maps, yet could display the necessary information in an easily understood format. The programs had to be written in such a way that they could be run with minimal operator training, but could still allow for a degree of customizing for scale and projection parameters to match various published and unpublished maps. That programming was accomplished through the use of AML in such a way that a single command line would produce a plotter-ready file. The use of a single command line allowed individuals with little or no GIS training to produce their own customized maps, and also allowed batch processing to proceed overnight when facilities are customarily underutilized. Programs were written to allow output to several printing devices, and to produce plots ranging from letter size to 36 inches wide by several feet long. The Geologic Division of the U. S. Geological Survey maintains an internet site where individuals may access and download public-domain data. Samples of some of the programs that were developed for the assessment have been made available at that site, and the address is given later in this discussion.

Since the purpose of an assessment is ultimately to present an estimate of resources, it is necessary to perform numerical calculations in addition to producing display maps. Maps showing production and exploration history can give a good sense of what has happened in the past, or what could happen in the future, but they give no numerical constraints on ultimate resources. In order to determine those constraints it is necessary to enter the realm of statisticians and economists. A GIS allows many statistical analyses to be performed using spatial data. Very simple calculations can be performed using the grid-cell data to determine the total area in which exploration has occurred, success rates, and potential area remaining to be explored. More intricate analyses required that programs had to be built to use GIS to sort through data files to find data that fell within specified plays. Those data were then used to build discovery-process models. AML was used to build programs that could run with a single command line, allowing large numbers of files to be accessed and sorted in a short period of time by a single user.

During the assessment, a series of programs were developed to build outlines of selected oil or gas fields. These programs were written in the form of a graphical user interface (GUI) that utilized menu bars and pulldown menus, allowing the user to point and click on the selected program and be prompted for the information necessary to complete the task. The original program was designed to automatically search the database for points labelled with a specific field name or a series of aliases, build a triangulated irregular network connecting the various points, dissolve the interior lines and put a 400-meter buffer around the resulting polygon. It was discovered that the information in the database was insufficient to completely automate the process. As a consequence, it was necessary to add a series of menus that would allow a geologist to edit the field outlines with the aid of digital background coverages and published state field maps. The resultant field outlines were used to select all the wells in the original database that fell within the field's boundary, and were also used to assign points from other databases to the same field. This allowed a one-to-one correspondence to be established between previously unconnected databases.

There are great advantages to having access to several databases and programs to access the data within them. There are also disadvantages when the data in different databases are in unrelated formats and accessed through different programs. The largest portion of the data available for the assessment was stored on a VAX system and utilized INGRES as the relational database management system (RDBMS). This data could be accessed using SQL or INGRES. The GIS work was done on a UNIX system with INFO, which is not a true RDBMS and does not support SQL, as a database management system. In order to avoid the necessity of duplicating entire databases constituting many gigabytes of storage, it was necessary to readily access the data across the network. This was accomplished through the use of the database integrator function in ARC/INFO. A copy of INGRES was loaded on the UNIX system that was also the host for ARC/INFO, and an interface package known as the Transmission Control Protocol/Internet Protocol (TCP/IP) was used to allow commands to be transmitted across the network to access the large databases. This "seamless connectivity" allowed the GIS to treat the large databases as if they were loaded on the local machine. The GIS was then utilized with SQL to perform complex data retrievals and calculations.

GIS was also used in the preparation of the data sets for final display purposes. A series of programs were written in AML to allow play-boundary lines and cell coverages to be drawn as graphics images on a McIntosh screen, where they were captured as bit maps. Those bit maps were then reproduced for display as a final product on the compact disk.

SUMMARY AND CONCLUSIONS

By necessity, the first steps in using GIS for assessment work consisted of building a database and refining the techniques for its utilization. With the basic structure now in place, ensuing assessment work can concentrate on developing more intricate programs to access and utilize data. Some future programs that could be developed might utilize GIS to define or assign areas of high or low risk within a particular play, map areas of source rocks or areas of similar reservoir characteristics, provide three-dimensional models of structures or reservoirs, map areas with a low potential for development due to environmental issues or population pressure, or map areas of high or low economic return. These and similar issues can be addressed through the utilization of GIS technology as scientists become more familiar with its capabilities.

Additional details of a more technical nature, including some representative programs, can be accessed through the Internet at the following address: greenwood@cr.usgs.gov (136.177.48.5). There is a public directory called "pub", and within that directory is a subdirectory called "beeman" where the sample programs reside.